NOVEL PROCOAGULANT PROTEINS

This invention relates to a novel series of proteins which exhibit procoagulant properties. These proteins have marked structural differences from human factor VIII:C, but have similar procoagulant activity.

Factor VIII:C is the blood plasma protein that is defective or absent in Hemophilia A disease. This disease is a hereditary bleeding disorder affecting approximately one in 20,000 males. The structure of factor VIII:C is described in U.S. Patent Applications Serial No. 546,650 filed October 28, 1983 and No. 644,036 filed August 24, 1984, which are incorporated herein by reference and in Nature, 312:306, 307, 326 and 342.

One of the problems presently encountered with the use of human factor VIII:C for treatment of hemophilia arises from its antigenicity. A significant percentage of hemophiliacs have developed an immune reaction to the factor VIII:C used for their treatment. Non-hemophiliacs can also develop or acquire hemophilia when their immune systems become sensitized to factor VIII:C and produce circulating antibodies or "inhibitors" to factor VIII:C. In either case, the effect is the neutralization of whatever factor VIII:C is present in the patient, making treatment very difficult. Until now, the method of choice for treating hemophiliacs with this problem has been to administer, in cases of severe bleeding episodes, non-human factor VIII:C, such as treated porcine factor VIII:C. See Kernoff et al., Blood 63:31 (1984). However, the antibodies which neutralize the clotting ability of human factor VIII:C will react to a varying extent with factor VIII:C of other species, and the porcine protein is itself antigenic, thus both the short-term and long-term effectiveness of such treatment will vary.

Additionally, patients frequently display adverse reactions to infusion with the porcine factor VIII:C. The use of porcine factor VIII:C in spite of the risks has been justified because of the lack of reliably effective alternatives. Kernoff, supra at 38. The present invention provides an alternative to the administration of porcine factor VIII:C.

This invention provides for proteins which have procoagulant activity similar to that of factor VIII:C and also have substantially lower molecular weight. These proteins are schematically depicted by formula (1) as follows:

(1) TI A-X-B TM (1)

wherein A represents a polypeptide sequence substantially duplicative of the sequence Ala-20 through Arg-759; B represents a polypeptide sequence substantially duplicative of the sequence Ser-1709 through the C-terminal Tyr-2351; and X represents a polypeptide sequence of up to 949 amino acids substantially duplicative of sequences of amino acids within the sequence Ser-760 through Arg-1708. The amino terminus of region X is covalently bonded through a peptide bond (designated "-" in formula 1) to the carboxy terminus of A. The carboxy terminus of region X is likewise bonded to the amino terminus of B. Numbering of amino acids throughout this disclosure is with reference to the numbering of amino acids in Table 1 in which the first amino acid, Met, of the leader sequence is assigned Number 1. Protein domain X may comprise a continuous but shorter sequence selected from the region Ser-760 through Arg-1708. Alternatively X may comprise two or more amino acid sequences selected from that region which are covalently bonded by a peptide bond (maintaining an ascending numerical order of amino acids).

By way of example, one compound of this invention contains a region X comprising the amino acid sequence of Ser-760 to Pro-

T0040X

TABLE 1

```
                5' GAATTCCCCAGTGGGTAAGTTCCTTAAATGCTCTGGAAAGAAATTCGGACTTTTCATTAAATCAGAAATT

TTACTTTTTTCCCCTCCTGCGAGCTAAAGATATTTTAGAGAAGAATTAACCTTTTGCTTCTCCAGTTGAACATTTCTAGCAATAAGTC

MET Gln Ile Glu Leu Ser Thr Cys Phe Phe Leu Cys Leu Leu Arg Phe Cys Phe  18
ATG CAA ATA GAG CTC TCC ACC TGC TTC TTT CTG TGC CTT TTG CGA TTC TGC TTT

Ser Ala Thr Arg Arg Tyr Tyr Leu Gly Ala Val Glu Leu Ser Trp Asp Tyr MET  36
AGT GCC ACC AGA AGA TAC TAC CTG GGT GCA GTG GAA CTG TCA TGG GAC TAT ATG

Gln Ser Asp Leu Gly Glu Leu Pro Val Asp Ala Arg Phe Pro Pro Arg Val Pro  54
CAA AGT GAT CTC GGT GAG CTG CCT GTG GAC GCA AGA TTT CCT CCT AGA GTG CCA

Lys Ser Phe Pro Phe Asn Thr Ser Val Val Tyr Lys Lys Thr Leu Phe Val Glu  72
AAA TCT TTT CCA TTC AAC ACC TCA GTC GTG TAC AAA AAG ACT CTG TTT GTA GAA

Phe Thr Val His Leu Phe Asn Ile Ala Lys Pro Arg Pro Pro Trp MET Gly Leu  90
TTC ACG GTT CAC CTT TTC AAC ATC GCT AAG CCA AGG CCA CCC TGG ATG GGT CTG

Leu Gly Pro Thr Ile Gln Ala Glu Val Tyr Asp Thr Val Val Ile Thr Leu Lys 108
CTA GGT CCT ACC ATC CAG GCT GAG GTT TAT GAT ACA GTG GTC ATT ACA CTT AAG

Asn MET Ala Ser His Pro Val Ser Leu His Ala Val Gly Val Ser Tyr Trp Lys 126
AAC ATG GCT TCC CAT CCT GTC AGT CTT CAT GCT GTT GGT GTA TCC TAC TGG AAA

Ala Ser Glu Gly Ala Glu Tyr Asp Asp Gln Thr Ser Gln Arg Glu Lys Glu Asp 144
GCT TCT GAG GGA GCT GAA TAT GAT GAT CAG ACC AGT CAA AGG GAG AAA GAA GAT

Asp Lys Val Phe Pro Gly Gly Ser His Thr Tyr Val Trp Gln Val Leu Lys Glu 162
GAT AAA GTC TTC CCT GGT GGA AGC CAT ACA TAT GTC TGG CAG GTC CTG AAA GAG

Asn Gly Pro MET Ala Ser Asp Pro Leu Cys Leu Thr Tyr Ser Tyr Leu Ser His 180
AAT GGT CCA ATG GCC TCT GAC CCA CTG TGC CTT ACC TAC TCA TAT CTT TCT CAT

Val Asp Leu Val Lys Asp Leu Asn Ser Gly Leu Ile Gly Ala Leu Leu Val Cys 198
GTG GAC CTG GTA AAA GAC TTG AAT TCA GGC CTC ATT GGA GCC CTA CTA GTA TGT

Arg Glu Gly Ser Leu Ala Lys Glu Lys Thr Gln Thr Leu His Lys Phe Ile Leu 216
AGA GAA GGG AGT CTG GCC AAG GAA AAG ACA CAG ACC TTG CAC AAA TTT ATA CTA

Leu Phe Ala Val Phe Asp Glu Gly Lys Ser Trp His Ser Glu Thr Lys Asn Ser 234
CTT TTT GCT GTA TTT GAT GAA GGG AAA AGT TGG CAC TCA GAA ACA AAG AAC TCC

Leu MET Gln Asp Arg Asp Ala Ala Ser Ala Arg Ala Trp Pro Lys MET His Thr 252
TTG ATG CAG GAT AGG GAT GCT GCA TCT GCT CGG GCC TGG CCT AAA ATG CAC ACA

Val Asn Gly Tyr Val Asn Arg Ser Leu Pro Gly Leu Ile Gly Cys His Arg Lys 270
GTC AAT GGT TAT GTA AAC AGG TCT CTG CCA GGT CTG ATT GGA TGC CAC AGG AAA

Ser Val Tyr Trp His Val Ile Gly MET Gly Thr Thr Pro Glu Val His Ser Ile 288
TCA GTC TAT TGG CAT GTG ATT GGA ATG GGC ACC ACT CCT GAA GTC CAC TCA ATA

Phe Lue Glu Gly His Thr Phe Leu Val Arg Asn His Arg Gln Ala Ser Leu Glu 306
TTC CTC GAA GGT CAC ACA TTT CTT GTG AGG AAC CAT CGC CAC GCG TCC TTG GAA
```

TABLE 1, continued

Ile	Ser	Pro	Ile	Thr	Phe	Leu	Thr	Ala	Gln	Thr	Leu	Leu	MET	Asp	Leu	Gly	Gln	324
ATC	TCG	CCA	ATA	ACT	TTC	CTT	ACT	GCT	CAA	ACA	CTC	TTG	ATG	GAC	CTT	GGA	CAG	
Phe	Leu	Leu	Phe	Cys	His	Ile	Ser	Ser	His	Gln	His	Asp	Gly	MET	Glu	Ala	Tyr	342
TTT	CTA	CTG	TTT	TGT	CAT	ATC	TCT	TCC	CAC	CAA	CAT	GAT	GGC	ATG	GAA	GCT	TAT	
Val	Lys	Val	Asp	Ser	Cys	Pro	Glu	Glu	Pro	Gln	Leu	Arg	MET	Lys	Asn	Asn	Glu	360
GTC	AAA	GTA	GAC	AGC	TGT	CCA	GAG	GAA	CCC	CAA	CTA	CGA	ATG	AAA	AAT	AAT	GAA	
Glu	Ala	Glu	Asp	Tyr	Asp	Asp	Asp	Leu	Thr	Asp	Ser	Glu	MET	Asp	Val	Val	Arg	378
GAA	GCG	GAA	GAC	TAT	GAT	GAT	GAT	CTT	ACT	GAT	TCT	GAA	ATG	GAT	GTG	GTC	AGG	
Phe	Asp	Asp	Asp	Asn	Ser	Pro	Ser	Phe	Ile	Gln	Ile	Arg	Ser	Val	Ala	Lys	Lys	396
TTT	GAT	GAT	GAC	AAC	TCT	CCT	TCC	TTT	ATC	CAA	ATT	CGC	TCA	GTT	GCG	AAG	AAG	
His	Pro	Lys	Thr	Trp	Val	His	Tyr	Ile	Asn	Ala	Glu	Glu	Glu	Asp	Trp	Asp	Tyr	414
CAT	CCT	AAA	ACT	TGG	GTA	CAT	TAC	ATT	GCT	GCT	GAA	GAG	GAG	GAC	TGG	GAC	TAT	
Ala	Pro	Leu	Val	Leu	Ala	Pro	Asp	Asp	Arg	Ser	Tyr	Lys	Ser	Gln	Tyr	Leu	Asn	432
GCT	CCC	TTA	GTC	CTC	GCC	CCC	GAT	GAC	AGA	AGT	TAT	AAA	AGT	CAA	TAT	TTG	AAC	
Asn	Gly	Pro	Gln	Arg	Ile	Gly	Arg	Lys	Tyr	Lys	Lys	Val	Arg	Phe	MET	Ala	Tyr	450
AAT	GGC	CCT	CAG	CGG	ATT	GGT	AGG	AAA	TAC	AAA	AAA	GTC	CGA	TTT	ATG	GCA	TAC	
Thr	Asp	Glu	Thr	Phe	Lys	Thr	Arg	Glu	Ala	Ile	Gln	His	Glu	Ser	Gly	Ile	Leu	468
ACA	GAT	GAA	ACC	TTT	AAG	ACT	CGT	GAA	GCT	ATT	CAG	CAT	GAA	TCA	GGA	ATC	TTG	
Gly	Pro	Leu	Leu	Tyr	Gly	Glu	Val	Gly	Asp	Thr	Leu	Leu	Ile	Ile	Phe	Lys	Asn	486
GGA	CCT	TTA	CTT	TAT	GGG	GAA	GTT	GGA	GAC	ACA	CTC	TTG	ATT	ATA	TTT	AAG	AAT	
Gln	Ala	Ser	Arg	Pro	Tyr	Asn	Ile	Tyr	Pro	His	Gly	Ile	Thr	Asp	Val	Arg	Pro	504
CAA	GCA	AGC	AGA	CCA	TAT	AAC	ATC	TAC	CCT	CAC	GGA	ATC	ACT	GAT	GTC	CGT	CCT	
Leu	Tyr	Ser	Arg	Arg	Leu	Pro	Lys	Gly	Val	Lys	His	Leu	Lys	Asp	Phe	Pro	Ile	522
TTG	TAT	TCA	AGG	AGA	TTA	CCA	AAA	GGT	GTA	AAA	CAT	TTG	AAG	GAT	TTT	CCA	ATT	
Leu	Pro	Gly	Glu	Ile	Phe	Lys	Tyr	Lys	Trp	Thr	Val	Thr	Val	Glu	Asp	Gly	Pro	540
CTG	CCA	GGA	GAA	ATA	TTC	AAA	TAT	AAA	TGG	ACA	GTG	ACT	GTA	GAA	GAT	GGG	CCA	
Thr	Lys	Ser	Asp	Pro	Arg	Cys	Leu	Thr	Arg	Tyr	Tyr	Ser	Ser	Phe	Val	Asn	MET	558
ACT	AAA	TCA	GAT	CCT	CGG	TGC	CTG	ACC	CGC	TAT	TAC	TCT	AGT	TTC	GTT	AAT	ATG	
Glu	Arg	Asp	Leu	Ala	Ser	Gly	Leu	Ile	Gly	Pro	Leu	Leu	Ile	Cys	Tyr	Lys	Glu	576
GAG	AGA	GAT	CTA	GCT	TCA	GGA	CTC	ATT	GGC	CCT	CTC	CTG	ATC	TGC	TAC	AAA	GAA	
Ser	Val	Asp	Gln	Arg	Gly	Asn	Gln	Ile	MET	Ser	Asp	Lys	Arg	Asn	Val	Ile	Leu	594
TCT	GTA	GAT	CAA	AGA	GGA	AAC	CAG	ATA	ATG	TCA	GAC	AAG	AGG	AAT	GTC	ATC	CTG	
Phe	Ser	Val	Phe	Asp	Glu	Asn	Arg	Ser	Trp	Tyr	Leu	Thr	Glu	Asn	Ile	Gln	Arg	612
TTT	TCT	GTA	TTT	GAT	GAG	AAC	CGA	AGC	TGG	TAC	CTC	ACA	GAG	AAT	ATA	CAA	CGC	
Phe	Leu	Pro	Asn	Pro	Ala	Gly	Val	Gln	Leu	Glu	Asp	Pro	Glu	Phe	Gln	Ala	Ser	630
TTT	CTC	CCC	AAT	CCA	GCT	GGA	GTG	CAG	CTT	GAG	GAT	CCA	GAG	TTC	CAA	GCC	TCC	
Asn	Ile	MET	His	Ser	Ile	Asn	Gly	Tyr	Val	Phe	Asp	Ser	Leu	Gln	Leu	Ser	Val	648
AAC	ATC	ATG	CAC	AGC	ATC	AAT	GGC	TAT	GTT	TTT	GAT	AGT	TTG	CAG	TTG	TCA	GTT	

TABLE 1, continued

Cys	Leu	His	Glu	Val	Ala	Tyr	Trp	Tyr	Ile	Leu	Ser	Ile	Gly	Ala	Gln	Thr	Asp	666
TGT	TTG	CAT	GAG	GTG	GCA	TAC	TGG	TAC	ATT	CTA	AGC	ATT	GGA	GCA	CAG	ACT	GAC	
Phe	Leu	Ser	Val	Phe	Phe	Ser	Gly	Tyr	Thr	Phe	Lys	His	Lys	MET	Val	Tyr	Glu	684
TTC	CTT	TCT	GTC	TTC	TTC	TCT	GGA	TAT	ACC	TTC	AAA	CAC	AAA	ATG	GTC	TAT	GAA	
Asp	Thr	Leu	Thr	Leu	Phe	Pro	Phe	Ser	Gly	Glu	Thr	Val	Phe	MET	Ser	MET	Glu	702
GAC	ACA	CTC	ACC	CTA	TTC	CCA	TTC	TCA	GGA	GAA	ACT	GTC	TTC	ATG	TCG	ATG	GAA	
Asn	Pro	Gly	Leu	Trp	Ile	Leu	Gly	Cys	His	Asn	Ser	Asp	Phe	Arg	Asn	Arg	Gly	720
AAC	CCA	GGT	CTA	TGG	ATT	CTG	GGG	TGC	CAC	AAC	TCA	GAC	TTT	CGG	AAC	AGA	GGC	
MET	Thr	Ala	Leu	Leu	Lys	Val	Ser	Ser	Cys	Asp	Lys	Asn	Thr	Gly	Asp	Tyr	Tyr	738
ATG	ACC	GCC	TTA	CTG	AAG	GTT	TCT	AGT	TGT	GAC	AAG	AAC	ACT	GGT	GAT	TAT	TAC	
Glu	Asp	Ser	Tyr	Glu	Asp	Ile	Ser	Ala	Tyr	Leu	Leu	Ser	Lys	Asn	Asn	Ala	Ile	756
GAG	GAC	AGT	TAT	GAA	GAT	ATT	TCA	GCA	TAC	TTG	CTG	AGT	AAA	AAC	AAT	GCC	ATT	
Glu	Pro	Arg	Ser	Phe	Ser	Gln	Asn	Ser	Arg	His	Pro	Ser	Thr	Arg	Gln	Lys	Gln	774
GAA	CCA	AGA	AGC	TTC	TCC	CAG	AAT	TCA	AGA	CAC	CCT	AGC	ACT	AGG	CAA	AAG	CAA	
Phe	Asn	Ala	Thr	Thr	Ile	Pro	Glu	Asn	Asp	Ile	Glu	Lys	Thr	Asp	Pro	Trp	Phe	792
TTT	AAT	GCC	ACC	ACA	ATT	CCA	GAA	AAT	GAC	ATA	GAG	AAG	ACT	GAC	CCT	TGG	TTT	
Ala	His	Arg	Thr	Pro	MET	Pro	Lys	Ile	Gln	Asn	Val	Ser	Ser	Ser	Asp	Leu	Leu	810
GCA	CAC	AGA	ACA	CCT	ATG	CCT	AAA	ATA	CAA	AAT	GTC	TCC	TCT	AGT	GAT	TTG	TTG	
MET	Leu	Leu	Arg	Gln	Ser	Pro	Thr	Pro	His	Gly	Leu	Ser	Leu	Ser	Asp	Leu	Gln	828
ATG	CTC	TTG	CGA	CAG	AGT	CCT	ACT	CCA	CAT	GGG	CTA	TCC	TTA	TCT	GAT	CTC	CAA	
Glu	Ala	Lys	Tyr	Glu	Thr	Phe	Ser	Asp	Asp	Pro	Ser	Pro	Gly	Ala	Ile	Asp	Ser	846
GAA	GCC	AAA	TAT	GAG	ACT	TTT	TCT	GAT	GAT	CCA	TCA	CCT	GGA	GCA	ATA	GAC	AGT	
Asn	Asn	Ser	Leu	Ser	Glu	MET	Thr	His	Phe	Arg	Pro	Gln	Leu	His	His	Ser	Gly	864
AAT	AAC	AGC	CTG	TCT	GAA	ATG	ACA	CAC	TTC	AGG	CCA	CAG	CTC	CAT	CAC	AGT	GGG	
Asp	MET	Val	Phe	Thr	Pro	Glu	Ser	Gly	Leu	Gln	Leu	Arg	Leu	Asn	Glu	Lys	Leu	882
GAC	ATG	GTA	TTT	ACC	CCT	GAG	TCA	GGC	CTC	CAA	TTA	AGA	TTA	AAT	GAG	AAA	CTG	
Gly	Thr	Thr	Ala	Ala	Thr	Glu	Leu	Lys	Lys	Leu	Asp	Phe	Lys	Val	Ser	Ser	Thr	900
GGC	ACA	ACT	GCA	GCA	ACA	GAG	TTG	AAG	AAA	CTT	GAT	TTC	AAA	GTT	TCT	AGT	ACA	
Ser	Asn	Asn	Leu	Ile	Ser	Thr	Ile	Pro	Ser	Asp	Asn	Leu	Ala	Ala	Gly	Thr	Asp	918
TCA	AAT	AAT	CTG	ATT	TCA	ACA	ATT	CCA	TCA	GAC	AAT	TTG	GCA	GCA	GGT	ACT	GAT	
Asn	Thr	Ser	Ser	Leu	Gly	Pro	Pro	Ser	MET	Pro	Val	His	Tyr	Asp	Ser	Gln	Leu	936
AAT	ACA	AGT	TCC	TTA	GGA	CCC	CCA	AGT	ATG	CCA	GTT	CAT	TAT	GAT	AGT	CAA	TTA	
Asp	Thr	Thr	Leu	Phe	Gly	Lys	Lys	Ser	Ser	Pro	Leu	Thr	Glu	Ser	Gly	Gly	Pro	954
GAT	ACC	ACT	CTA	TTT	GGC	AAA	AAG	TCA	TCT	CCC	CTT	ACT	GAG	TCT	GGT	GGA	CCT	
Leu	Ser	Leu	Ser	Glu	Glu	Asn	Asn	Asp	Ser	Lys	Leu	Leu	Glu	Ser	Gly	Leu	MET	972
CTC	AGC	TTG	AGT	GAA	GAA	AAT	AAT	GAT	TCA	AAG	TTG	TTA	GAA	TCA	GGT	TTA	ATG	
Asn	Ser	Gln	Glu	Ser	Ser	Trp	Gly	Lys	Asn	Val	Ser	Ser	Thr	Glu	Ser	Gly	Arg	990
AAT	AGC	CAA	GAA	AGT	TCA	TGG	GGA	AAA	AAT	GTA	TCC	TCA	ACA	GAG	AGT	GGT	AGG	

TABLE 1, continued

Leu	Phe	Lys	Gly	Lys	Arg	Ala	His	Gly	Pro	Ala	Leu	Leu	Thr	Lys	Asp	Asn	Ala	1,008
TTA	TTT	AAA	GGC	AAA	AGA	GCT	CAT	GGA	CCT	GCT	TTG	TTG	ACT	AAA	GAT	AAT	GCC	
Leu	Phe	Lys	Val	Ser	Ile	Ser	Leu	Leu	Lys	Thr	Asn	Lys	Thr	Ser	Asn	Asn	Ser	1,026
TTA	TTC	AAA	GTT	AGC	ATC	TCT	TTG	TTA	AAG	ACA	AAC	AAA	ACT	TCC	AAT	AAT	TCA	
Ala	Thr	Asn	Arg	Lys	Thr	His	Ile	Asp	Gly	Pro	Ser	Leu	Leu	Ile	Glu	Asn	Ser	1,044
GCA	ACT	AAT	AGA	AAG	ACT	CAC	ATT	GAT	GGC	CCA	TCA	TTA	TTA	ATT	GAG	AAT	AGT	
Pro	Ser	Val	Trp	Gln	Asn	Ile	Leu	Glu	Ser	Asp	Thr	Glu	Phe	Lys	Lys	Val	Thr	1,062
CCA	TCA	GTC	TGG	CAA	AAT	ATA	TTA	GAA	AGT	GAC	ACT	GAG	TTT	AAA	AAA	GTG	ACA	
Pro	Leu	Ile	His	Asp	Arg	MET	Leu	MET	Asp	Lys	Asn	Ala	Thr	Ala	Leu	Arg	Leu	1,080
CCT	TTG	ATT	CAT	GAC	AGA	ATG	CTT	ATG	GAC	AAA	AAT	GCT	ACA	GCT	TTG	AGG	CTA	
Asn	His	MET	Ser	Asn	Lys	Thr	Thr	Ser	Ser	Lys	Asn	MET	Glu	MET	Val	Gln	Gln	1,098
AAT	CAT	ATG	TCA	AAT	AAA	ACT	ACT	TCA	TCA	AAA	ACC	ATG	GAA	ATG	GTC	CAA	CAG	
Lys	Lys	Glu	Gly	Pro	Ile	Pro	Pro	Asp	Ala	Gln	Asn	Pro	Asp	MET	Ser	Phe	Phe	1,116
AAA	AAA	GAG	GGC	CCC	ATT	CCA	CCA	GAT	GCA	CAA	AAT	CCA	GAT	ATG	TCG	TTC	TTT	
Lys	MET	Leu	Phe	Leu	Pro	Glu	Ser	Ala	Arg	Trp	Ile	Gln	Arg	Thr	His	Gly	Lys	1,134
AAG	ATG	CTA	TTC	TTG	CCA	GAA	TCA	GCA	AGG	TGG	ATA	CAA	AGG	ACT	CAT	GGA	AAG	
Asn	Ser	Leu	Asn	Ser	Gly	Gln	Gly	Pro	Ser	Pro	Lys	Gln	Leu	Val	Ser	Leu	Gly	1,152
AAC	TCT	CTG	AAC	TCT	GGG	CAA	GGC	CCC	AGT	CCA	AAG	CAA	TTA	GTA	TCC	TTA	GGA	
Pro	Glu	Lys	Ser	Val	Glu	Gly	Gln	Asn	Phe	Leu	Ser	Glu	Lys	Asn	Lys	Val	Val	1,170
CCA	GAA	AAA	TCT	GTG	GAA	GGT	CAG	AAT	TTC	TTG	TCT	GAG	AAA	AAC	AAA	GTC	GTA	
Val	Gly	Lys	Gly	Glu	Phe	Thr	Lys	Asp	Val	Gly	Leu	Lys	Glu	MET	Val	Phe	Pro	1,188
GTA	GGA	AAG	GGT	GAA	TTT	ACA	AAG	GAC	GTA	GGA	CTC	AAA	GAG	ATG	GTT	TTT	CCA	
Ser	Ser	Arg	Asn	Leu	Phe	Leu	Thr	Asn	Leu	Asp	Asn	Leu	His	Glu	Asn	Asn	Thr	1,206
AGC	AGC	AGA	AAC	CTA	TTT	CTT	ACT	AAC	TTG	GAT	AAT	TTA	CAT	GAA	AAT	AAT	ACA	
His	Asn	Gln	Glu	Lys	Lys	Ile	Gln	Glu	Glu	Ile	Glu	Lys	Lys	Glu	Thr	Leu	Ile	1,224
CAC	AAT	CAA	GAA	AAA	AAA	ATT	CAG	GAA	GAA	ATA	GAA	AAG	AAG	GAA	ACA	TTA	ATC	
Gln	Glu	Asn	Val	Val	Leu	Pro	Gln	Ile	His	Thr	Val	Thr	Gly	Thr	Lys	Asn	Phe	1,242
CAA	GAG	AAT	GTA	GTT	TTG	CCT	CAG	ATA	CAT	ACA	GTG	ACT	GGC	ACT	AAG	AAT	TTC	
MET	Lys	Asn	Leu	Phe	Leu	Leu	Ser	Thr	Arg	Gln	Asn	Val	Glu	Gly	Ser	Tyr	Glu	1,260
ATG	AAG	AAC	CTT	TTC	TAA	CTG	AGC	ACT	AGG	CAA	AAT	GTA	GAA	GGT	TCA	TAT	GAG	
Gly	Ala	Tyr	Ala	Pro	Val	Leu	Gln	Asp	Phe	Arg	Ser	Leu	Asn	Asp	Ser	Thr	Asn	1,278
GGC	GCA	TAT	GCT	CCA	GTA	CTT	CAA	GAT	TTT	AGG	TCA	TTA	AAT	GAT	TCA	ACA	AAT	
Arg	Thr	Lys	Lys	His	Thr	Ala	His	Phe	Ser	Lys	Lys	Gly	Glu	Glu	Glu	Asn	Leu	1,296
AGA	ACA	AAG	AAA	CAC	ACA	GCT	CAT	TTC	TCA	AAA	AAA	GGC	GAG	GAA	GAA	AAC	TTG	
Glu	Gly	Leu	Gly	Asn	Gln	Thr	Lys	Gln	Ile	Val	Glu	Lys	Tyr	Ala	Cys	Thr	Thr	1,314
GAA	GGC	TTG	GGA	AAT	CAA	ACC	AAG	CAA	ATT	GTA	GAG	AAA	TAT	GCA	TGC	ACC	ACA	
Arg	Ile	Ser	Pro	Asn	Thr	Ser	Gln	Gln	Asn	Phe	Val	Thr	Gln	Arg	Ser	Lys	Arg	1,332
AGG	ATA	TCT	CCT	AAT	ACA	AGC	CAG	CAG	AAT	TTT	GTC	ACG	CAA	CGT	AGT	AAG	AGA	

TABLE 1, continued

Ala	Leu	Lys	Gln	Phe	Arg	Leu	Pro	Leu	Glu	Glu	Thr	Glu	Leu	Glu	Lys	Arg	Ile	1,350
GCT	TTG	AAA	CAA	TTC	AGA	CTC	CCA	CTA	GAA	GAA	ACA	GAA	CTT	GAA	AAA	AGG	ATA	
Ile	Val	Asp	Asp	Thr	Ser	Thr	Gln	Trp	Ser	Lys	Asn	MET	Lys	His	Leu	Thr	Pro	1,368
ATT	GTG	GAT	GAC	ACC	TCA	ACC	CAG	TGG	TCC	AAA	AAC	ATG	AAA	CAT	TTG	ACC	CCG	
Ser	Thr	Leu	Thr	Gln	Ile	Asp	Tyr	Asn	Glu	Lys	Glu	Lys	Gly	Ala	Ile	Thr	Gln	1,386
AGC	ACC	CTC	ACA	CAG	ATA	GAC	TAC	AAT	GAG	AAG	GAG	AAA	GGG	GCC	ATT	ACT	CAG	
Ser	Pro	Leu	Ser	Asp	Cys	Leu	Thr	Arg	Ser	His	Ser	Ile	Pro	Gln	Ala	Asn	Arg	1,404
TCT	CCC	TTA	TCA	GAT	TGC	CTT	ACG	AGG	AGT	CAT	AGC	ATC	CCT	CAA	GCA	AAT	AGA	
Ser	Pro	Leu	Pro	Ile	Ala	Lys	Val	Ser	Ser	Phe	Pro	Ser	Ile	Arg	Pro	Ile	Tyr	1,422
TCT	CCA	TTA	CCC	ATT	GCA	AAG	GTA	TCA	TCA	TTT	CCA	TCT	ATT	AGA	CCT	ATA	TAT	
Leu	Thr	Arg	Val	Leu	Phe	Gln	Asp	Asn	Ser	Ser	His	Leu	Pro	Ala	Ala	Ser	Tyr	1,440
CTG	ACC	AGG	GTC	CTA	TTC	CAA	GAC	AAC	TCT	TCT	CAT	CTT	CCA	GCA	GCA	TCT	TAT	
Arg	Lys	Lys	Asp	Ser	Gly	Val	Gln	Glu	Ser	Ser	His	Phe	Leu	Gln	Gly	Ala	Lys	1,458
AGA	AAG	AAA	GAT	TCT	GGG	GTC	CAA	GAA	AGC	AGT	CAT	TTC	TTA	CAA	GGA	GCC	AAA	
Lys	Asn	Asn	Leu	Ser	Leu	Ala	Ile	Leu	Thr	Leu	Glu	MET	Thr	Gly	Asp	Gln	Arg	1,476
AAA	AAT	AAC	CTT	TCT	TTA	GCC	ATT	CTA	ACC	TTG	GAG	ATG	ACT	GGT	GAT	CAA	AGA	
Glu	Val	Gly	Ser	Leu	Gly	Thr	Ser	Ala	Thr	Asn	Ser	Val	Thr	Tyr	Lys	Lys	Val	1,494
GAG	GTT	GGC	TCC	CTG	GGG	ACA	AGT	GCC	ACA	AAT	TCA	GTC	ACA	TAC	AAG	AAA	GTT	
Glu	Asn	Thr	Val	Leu	Pro	Lys	Pro	Asp	Leu	Pro	Lys	Thr	Ser	Gly	Lys	Val	Glu	1,512
GAG	AAC	ACT	GTT	CTC	CCG	AAA	CCA	GAC	TTG	CCC	AAA	ACA	TCT	GGC	AAA	GTT	GAA	
Leu	Leu	Pro	Lys	Val	His	Ile	Tyr	Gln	Lys	Asp	Leu	Phe	Pro	Thr	Glu	Thr	Ser	1,530
TTG	CTT	CCA	AAA	GTT	CAC	ATT	TAT	CAG	AAG	GAC	CTA	TTC	CCT	ACG	GAA	ACT	AGC	
Asn	Gly	Ser	Pro	Gly	His	Leu	Asp	Leu	Val	Glu	Gly	Ser	Leu	Leu	Gln	Gly	Thr	1,548
AAT	GGG	TCT	CCT	GGC	CAT	CTG	GAT	CTC	GTG	GAA	GGG	AGC	CTT	CTT	CAG	GGA	ACA	
Glu	Gly	Ala	Ile	Lys	Trp	Asn	Glu	Ala	Asn	Arg	Pro	Gly	Lys	Val	Pro	Phe	Leu	1,566
GAG	GGA	GCG	ATT	AAG	TGG	AAT	GAA	GCA	AAC	AGA	CCT	GGA	AAA	GTT	CCC	TTT	CTG	
Arg	Val	Ala	Thr	Glu	Ser	Ser	Ala	Lys	Thr	Pro	Ser	Lys	Leu	Leu	Asp	Pro	Leu	1,584
AGA	GTA	GCA	ACA	GAA	AGC	TCT	GCA	AAG	ACT	CCC	TCC	AAG	CTA	TTG	GAT	CCT	CTT	
Ala	Trp	Asp	Asn	His	Tyr	Gly	Thr	Gln	Ile	Pro	Lys	Glu	Glu	Trp	Lys	Ser	Gln	1,602
GCT	TGG	GAT	AAC	CAC	TAT	GGT	ACT	CAG	ATA	CCA	AAA	GAA	GAG	TGG	AAA	TCC	CAA	
Glu	Lys	Ser	Pro	Glu	Lys	Thr	Ala	Phe	Lys	Lys	Lys	Asp	Thr	Ile	Leu	Ser	Leu	1,520
GAG	AAG	TCA	CCA	GAA	AAA	ACA	GCT	TTT	AAG	AAA	AAG	GAT	ACC	ATT	TTG	TCC	CTG	
Asn	Ala	Cys	Glu	Ser	Asn	His	Ala	Ile	Ala	Ala	Ile	Asn	Glu	Gly	Gln	Asn	Lys	1,638
AAC	GCT	TGT	GAA	AGC	AAT	CAT	GCA	ATA	GCA	GCA	ATA	AAT	GAG	GGA	CAA	AAT	AAG	
Pro	Glu	Ile	Glu	Val	Thr	Trp	Ala	Lys	Gln	Gly	Arg	Thr	Glu	Arg	Leu	Cys	Ser	1,656
CCC	GAA	ATA	GAA	GTC	ACC	TGG	GCA	AAG	CAA	GGT	AGG	ACT	GAA	AGG	CTC	TGC	TCT	
Gln	Asn	Pro	Pro	Val	Leu	Lys	Arg	His	Gln	Arg	Glu	Ile	Thr	Arg	Thr	Thr	Leu	1,674
CAA	AAC	CCA	CCA	GTC	TTG	AAA	CGC	CAT	CAA	CGC	GAA	ATA	ACT	CGT	ACT	ACT	CTT	

TABLE 1, continued

Gln	Ser	Asp	Gln	Glu	Glu	Ile	Asp	Tyr	Asp	Asp	Thr	Ile	Ser	Val	Glu	MET	Lys	1,692
CAG	TCA	GAT	CAA	GAG	GAA	ATT	GAC	TAT	GAT	GAT	ACC	ATA	TCA	GTT	GAA	ATG	AAG	
Lys	Glu	Asp	Phe	Asp	Ile	Tyr	Asp	Glu	Asp	Glu	Asn	Gln	Ser	Pro	Arg	Ser	Phe	1,710
AAG	GAA	GAT	TTT	GAC	ATT	TAT	GAT	GAG	GAT	GAA	AAT	CAG	AGC	CCC	CGC	AGC	TTT	
Gln	Lys	Lys	Thr	Arg	His	Tyr	Phe	Ile	Ala	Ala	Val	Glu	Arg	Leu	Trp	Asp	Tyr	1,728
CAA	AAG	AAA	ACA	CGA	CAC	TAT	TTT	ATT	GCT	GCA	GTC	GAG	AGG	CTC	TGG	GAT	TAT	
Gly	MET	Ser	Ser	Ser	Pro	His	Val	Leu	Arg	Asn	Arg	Ala	Gln	Ser	Gly	Ser	Val	1,746
GGC	ATG	AGT	AGC	TCC	CCA	CAT	GTT	CTA	AGA	AAC	AGG	GCT	CAG	AGT	GGC	AGT	GTC	
Pro	Gln	Phe	Lys	Lys	Val	Val	Phe	Gln	Glu	Phe	Thr	Asp	Gly	Ser	Phe	Thr	Gln	1,764
CCT	CAG	TTC	AAG	AAA	GTT	GTT	TTC	CAG	GAA	TTT	ACT	GAT	GGC	TCC	TTT	ACT	CAG	
Pro	Leu	Tyr	Arg	Gly	Glu	Leu	Asn	Glu	His	Leu	Gly	Leu	Leu	Gly	Pro	Tyr	Ile	1,782
CCC	TTA	TAC	CGT	GGA	GAA	CTA	AAT	GAA	CAT	TTG	GGA	CTC	CTG	GGC	CCA	TAT	ATA	
Arg	Ala	Glu	Val	Glu	Asp	Asn	Ile	MET	Val	Thr	Phe	Arg	Asn	Gln	Ala	Ser	Arg	1,800
AGA	GCA	GAA	GTT	GAA	GAT	AAT	ATC	ATG	GTA	ACT	TTC	AGA	AAT	CAG	GCC	TCT	CGT	
Pro	Tyr	Ser	Phe	Tyr	Ser	Ser	Leu	Ile	Ser	Tyr	Glu	Glu	Asp	Gln	Arg	Gln	Gly	1,818
CCC	TAT	TCC	TTC	TAT	TCT	AGC	CTT	ATT	TCT	TAT	GAG	GAA	GAT	CAG	AGG	CAA	GGA	
Ala	Glu	Pro	Arg	Lys	Asn	Phe	Val	Lys	Pro	Asn	Glu	Thr	Lys	Thr	Tyr	Phe	Trp	1,836
GCA	GAA	CCT	AGA	AAA	AAC	TTT	GTC	AAG	CCT	AAT	GAA	ACC	AAA	ACT	TAC	TTT	TGG	
Lys	Val	Gln	His	His	MET	Ala	Pro	Thr	Lys	Asp	Glu	Phe	Asp	cys	Lys	Ala	Trp	1,854
AAA	GTG	CAA	CAT	CAT	ATG	GCA	CCC	ACT	AAA	GAT	GAG	TTT	GAC	TGC	AAA	GCC	TGG	
Ala	Tyr	Phe	Ser	Asp	Val	Asp	Leu	Glu	Lys	Asp	Val	His	Ser	Gly	Leu	Ile	Gly	1,872
GCT	TAT	TTC	TCT	GAT	GTT	GAC	CTG	GAA	AAA	GAT	GTG	CAC	TCA	GGC	CTG	ATT	GGA	
Pro	Leu	Leu	Val	Cys	His	Thr	Asn	Thr	Leu	Asn	Pro	Ala	His	Gly	Arg	Gln	Val	1,890
CCC	CTT	CTG	CTC	TGC	CAC	ACT	AAC	ACA	CTG	AAC	CCT	GCT	CAT	GGG	AGA	CAA	GTG	
Thr	Val	Glu	Glu	Phe	Ala	Leu	Phe	Phe	Thr	Ile	Phe	Asp	Glu	Thr	Lys	Ser	Trp	1,908
ACA	GTA	CAG	GAA	TTT	GCT	CTG	TTT	TTC	ACC	ATC	TTT	GAT	GAG	ACC	AAA	AGC	TGG	
Thy	Phe	Thr	Glu	Asn	MET	Glu	Arg	Asn	Cys	Arg	Ala	Pro	Cys	Asn	Ile	Gln	MET	1,926
TAC	TTC	ACT	GAA	AAT	ATG	GAA	AGA	AAC	TGC	AGG	GCT	CCC	TGC	AAT	ATC	CAG	ATG	
Glu	Asp	Pro	Thr	Phe	Lys	Glu	Asn	Thr	Arg	Phe	His	Ala	Ile	Asn	Gly	Tyr	Ile	1,944
GAA	GAT	CCC	ACT	TTT	AAA	GAG	AAT	TAT	CGC	TTC	CAT	GCA	ATC	AAT	GGC	TAC	ATA	
MET	Asp	Thr	Leu	Pro	Gly	Leu	Val	MET	Ala	Gln	Asp	Gln	Arg	Ile	Arg	Trp	Tyr	1,962
ATG	GAT	ACA	CTA	CCT	GGC	TTA	GTA	ATG	GCT	CAG	GAT	CAA	AGG	ATT	CGA	TGG	TAT	
Leu	Leu	Ser	MET	Gly	Ser	Asn	Glu	Asn	Ile	His	Ser	Ile	His	Phe	Ser	Gly	His	1,980
CTC	CTC	AGC	ATG	GGC	AGC	AAT	GAA	AAC	ATC	CAT	TCT	ATT	CAT	TTC	AGT	GGA	CAT	
Val	Phe	Thr	Val	Arg	Lys	Lys	Glu	Glu	Tyr	Lys	MET	Ala	Leu	Tyr	Asn	Leu	Tyr	1,998
GTG	TTC	ACT	GTA	CGA	AAA	AAA	GAG	GAG	TAT	AAA	ATG	GCA	CTG	TAC	AAT	CTC	TAT	
Pro	Gly	Val	Phe	Glu	Thr	Val	Glu	MET	Leu	Pro	Ser	Lys	Ala	Gly	Ile	Trp	Arg	2,016
CCA	GGT	GTT	TTT	GAG	ACA	GTG	GAA	ATG	TTA	CCA	TCC	AAA	GCT	GGA	ATT	TGG	CGC	

TABLE 1, continued

Val	Glu	Cys	Leu	Ile	Gly	Glu	His	Leu	His	Ala	Gly	MET	Ser	Thr	Leu	Phe	Leu	2,034
GTC	GAA	TGC	CTT	ATT	GGC	GAG	CAT	CTA	CAT	GCT	GGG	ATG	AGC	ACA	CTT	TTT	CTG	
Val	Tyr	Ser	Asn	Lys	Cys	Glu	Thr	Pro	Leu	Gly	MET	Ala	Ser	Gly	His	Ile	Arg	2,052
GTG	TAC	AGC	AAT	AAG	TGT	GAG	ACT	CCC	CTG	GGA	ATG	GCT	TCT	GGA	CAC	ATT	AGA	
Asp	Phe	Gln	Ile	Thr	Ala	Ser	Gly	Gln	Tyr	Gly	Gln	Trp	Ala	Pro	Lys	Leu	Ala	2,070
GAT	TTT	CAG	ATT	ACA	GCT	TCA	GGA	CAA	TAT	GGA	CAG	TGG	GCC	CCA	AAG	CTG	GCC	
Arg	Leu	His	Tyr	Ser	Gly	Ser	Ile	Asn	Ala	Trp	Ser	Thr	Lys	Glu	Pro	Phe	Ser	2,088
AGA	CTT	CAT	TAT	TCC	GGA	TCA	ATC	AAT	GCC	TGG	AGC	ACC	AAG	GAG	CCC	TTT	TCT	
Trp	Ile	Lys	Val	Asp	Leu	Leu	Ala	Pro	MET	Ile	Ile	His	Gly	Ile	Lys	Thr	Gln	2,106
TGG	ATC	AAG	GTG	GAT	CTG	TTG	GCA	CCA	ATG	ATT	ATT	CAC	GGC	ATC	AAG	ACC	CAG	
Gly	Ala	Arg	Gln	Lys	Phe	Ser	Ser	Leu	Tyr	Ile	Ser	Gln	Phe	Ile	Ile	MET	Tyr	2,124
GGT	GCC	CGT	CAG	AAG	TTC	TCC	AGC	CTC	TAC	ATC	TCT	CAG	TTT	ATC	ATC	ATG	TAT	
Ser	Leu	Asp	Gly	Lys	Lys	Trp	Gln	Thr	Tyr	Arg	Gly	Asn	Ser	Thr	Gly	Thr	Leu	2,142
AGT	CTT	GAT	GGG	AAG	AAG	TGG	CAG	ACT	TAT	CGA	GGA	AAT	TCC	ACT	GGA	ACC	TTA	
MET	Val	Phe	Phe	Gly	Asn	Val	Asp	Ser	Ser	Gly	Ile	Lys	His	Asn	Ile	Phe	Asn	2,160
ATG	GTC	TTC	TTT	GGC	AAT	GTC	GAT	TCA	TCT	GGG	ATA	AAA	CAC	AAT	ATT	TTT	AAC	
Pro	Pro	Ile	Ile	Ala	Arg	Tyr	Ile	Arg	Leu	His	Pro	Thr	His	Tyr	Ser	Ile	Arg	2,178
CCT	CCA	ATT	ATT	GCT	CGA	TAC	ATC	CGT	TTG	CAC	CCA	ACT	CAT	TAT	AGC	ATT	CGC	
Ser	Thr	Leu	Arg	MET	Glu	Leu	MET	Gly	Cys	Asp	Leu	Asn	Ser	Cys	Ser	MET	Pro	2,196
AGC	ACT	CTT	CGC	ATG	GAG	TTG	ATG	GGC	TGT	GAT	TTA	AAT	AGT	TGC	AGC	ATG	CCA	
Leu	Gly	MET	Glu	Ser	Lys	Ala	Ile	Ser	Asp	Ala	Gln	Ile	Thr	Ala	Ser	Ser	Tyr	2,214
TTG	GGA	ATG	GAG	AGT	AAA	GCA	ATA	TCA	GAT	GCA	CAG	ATT	ACT	GCT	TCA	TCC	TAC	
Phe	Thr	Asn	MET	Phe	Ala	Thr	Trp	Ser	Pro	Ser	Lys	Ala	Arg	Leu	His	Leu	Gln	2,232
TTT	ACC	AAT	ATG	TTT	GCC	ACC	TGG	TCT	CCT	TCA	AAA	GCT	CGA	CTT	CAC	CTC	CAA	
Gly	Arg	Ser	Asn	Ala	Trp	Arg	Pro	Gln	Val	Asn	Asn	Pro	Lys	Glu	Trp	Leu	Gln	2,250
GGG	AGG	AGT	AAT	GCC	TGG	AGA	CCT	CAG	GTC	AAT	AAT	CCA	AAA	GAG	TGG	CTG	CAA	
Val	Asp	Phe	Gln	Lys	Thr	MET	Lys	Val	Thr	Gly	Val	Thr	Thr	Gln	Gly	Val	Lys	2,268
GTG	GAC	TTC	CAG	AAG	ACA	ATG	AAA	GTC	ACA	GGA	GTA	ACT	ACT	CAG	GGA	GTA	AAA	
Ser	Leu	Leu	Thr	Ser	MET	Tyr	Val	Lys	Glu	Phe	Leu	Ile	Ser	Ser	Ser	Gln	Asp	2,286
TCT	CTG	CTT	ACC	ACC	ATG	TAT	GTG	AAG	GAG	TTC	CTC	ATC	TCC	AGC	AGT	CAA	GAT	
Gly	His	Gln	Trp	Thr	Leu	Phe	Phe	Gln	Asn	Gly	Lys	Val	Lys	Val	Phe	Gln	Gly	2,304
GGC	CAT	CAG	TGG	ACT	CTC	TTT	TTT	CAG	AAT	GGC	AAA	GTA	AAG	GTT	TTT	CAG	GGA	
Asn	Gln	Asp	Ser	Phe	Thr	Pro	Val	Val	Asn	Ser	Leu	Asp	Pro	Pro	Leu	Leu	Thr	2,322
AAT	CAA	GAC	TCC	TTC	ACA	CCT	GTG	GTG	AAC	TCT	CTA	GAC	CCA	CCG	TTA	CTG	ACT	
Arg	Tyr	Leu	Arg	Ile	His	Pro	Gln	Ser	Trp	Val	His	Gln	Ile	Ala	Leu	Arg	MET	2,340
CGC	TAC	CTT	CGA	ATT	CAC	CCC	CAG	AGT	TGG	GTG	CAC	CAG	ATT	GCC	CTG	AGG	ATG	
Glu	Val	Leu	Gly	Cys	Glu	Ala	Gln	Asp	Leu	Tyr	End							2,352
GAG	GTT	CTG	GGC	TGC	GAG	GCA	CAG	GAC	CTC	TAC	TGA	GGGTGGCCACTGCATCCCACCTGCCAGTG						

CCGTCACCTCTCCCTCCTCAGCTCCAGGGCATGTGTCCCTCCCTGGCTTGCTTCTACCTTTGTGCTAAATCCTAGCAGACACTCCCTTG

AAGCCTCCTGAATTAACTATCATCAGTCCTGCATTTCTTTGGTGGGGGCCCAGGAGGCTGCATCCATTTTAACTTAACTCTTACCTATT

TTCTCCAGCTGCTCCCAGA

XI

1000 followed by the amino acid sequence of Asp-1582 to Arg-1708. That compound thus comprises the polypeptide sequence of Ala-20 to Pro-1000 covalently linked by a peptide bond to amino acids Asp-1582 to Tyr-2351. Another exemplary compound contains a region X comprising the amino acid sequence Ser-760 to Thr-778 followed by the sequence Pro-1659 to Arg-1708. That compound thus comprises the polypeptide sequence Ala-20 to Thr-778 covalently linked by a peptide bond to the sequence Pro-1659 through Tyr-2351. Still another exemplary compound contains a region X comprising the amino acid sequence Ser-760 to Thr-778 followed by the sequence Glu-1694 to Arg-1708. That compound thus comprises the polypeptide sequence Ala-20 to Thr-778 covalently linked by a peptide bond to amino acids Glu-1694 through Tyr-2351.

These exemplary compounds are depicted schematically in Table 2.

The amino acid sequence represented by X should be selected so that it does not substantially reduce the procoagulant activity of the molecule, which activity can be conveniently assayed by conventional methods. Compound (2) of Table 2 is a presently preferred embodiment.

The procoagulant protein may be produced by appropriate host cells transformed by factor VIII:C DNA which has been specifically altered by use of any of a variety of site-specific mutagenesis techniques which will be familiar to those of ordinary skill in the art of recombinant DNA.

The starting materials may be a DNA sequence which codes for the complete factor VIII:C molecule, e.g., the complete human factor VIII:C as shown in Table 1, a truncated version of that sequence, or it may comprise segments of that DNA sequence, so long as the starting materials contain at least sufficient DNA to code for the amino acid sequences of the desired polypeptide.

TABLE 2: EXEMPLARY COMPOUNDS A-X-B

Compound	Amino Acid Sequence	X	Deletion
(human factor VIII:c)	$(Ala_{20} \rightarrow Tyr_{2351})$	$(Ser_{760} \rightarrow Arg_{1708})$	0
1	$(Ala_{20} \rightarrow Pro_{1000})$-$(Asp_{1582} \rightarrow Tyr_{2351})$	$(Ser_{760} \rightarrow Pro_{1000})$-$(Asp_{1582} \rightarrow Arg_{1708})$	581
2	$(Ala_{20} \rightarrow Thr_{778})$ -$(Pro_{1659} \rightarrow Tyr_{2351})$	$(Ser_{760} \rightarrow Thr_{778})$ -$(Pro_{1659} \rightarrow Arg_{1708})$	880
3	$(Ala_{20} \rightarrow Thr_{778})$ -$(Glu_{1694} \rightarrow Tyr_{2351})$	$(Ser_{760} \rightarrow Thr_{778})$ -$(Glu_{1694} \rightarrow Arg_{1708})$	915

A and B are as defined, supra; "-" represents a peptide bond; "→" indicates a polypeptide sequence inclusive of the specified amino acids; amino acid numbering corresponds to the numbering of the sequence depicted in Table 1; and "deletion" indicates the number of amino acids deleted relative to human factor VIII:c.

T0120X

12

The ~~procoagulent~~ procoagulant proteins of the present invention, in addition to lacking a substantial amino acid segment of human factor VIII:C, also have fewer potential N-glycosylation sites than human factor VIII. Preferably, at least one N-glycosylation site has been deleted. More preferably, 18 of the 25 potential N-glycosylation sites are not in the molecule. In still more preferred embodiments, up to 19 of the 25 potential N-glycosylation sites are removed. While not wishing to be bound by theory, it is presently believed that the antibodies to factor VIII:C which are directed to antigenic determinants contained in the protein segment deleted in accordance with this invention, i.e., in the amino acid segement itself or in the carbohydrate portion of the glycosylated protein, will not neutralize the procoagulant proteins of the present invention. Moreover, the fact that the procoagulants of the present invention lack many of the sites for non-human glycosylation by the non-human mammalian or other cells used to produce the proteins is also belived to reduce the antigenicity of that protein, and lessen the likelihood of developing antibodies to the procoagulants. This may enable facilitating the treatment of patients in need of procoagulant therapy.

I contemplate that my compounds can be produced by recombinant DNA techniques at a much lower cost than is possible for production of human factor VIII. The host organisms should more efficiently process and express the substantially simpler molecules of this invention.

The compounds of this invention can be formulated into pharmaceutically acceptable preparations with parenterally acceptable vehicles and excipients in accordance with procedures known in the art.

The pharmaceutical preparations of this invention, suitable for parenteral administration, may conveniently comprise a sterile lyophilized preparation of the protein which may be reconsti-

tuted by addition of sterile solution to produce solutions preferably isotonic with the blood of the recipient. The preparation may be presented in unit or multi-dose containers, e.g. in sealed ampoules or vials. Their use would be analogous to that of human factor VIII, appropriately adjusted for potency.

One method by which these proteins can be expressed is by use of DNA which is prepared by cutting a full-length factor VIII:C DNA with the appropriate restriction enzymes to remove a portion of the DNA sequence that codes for amino acids 760 to 1708 of human factor VIII:C. The cut DNA is then ligated with an oligonucleotide that resects the cut DNA and maintains the correct translational reading frame.

Preparation of the cDNA has been set forth in detail in U.S. Patent Applications Serial Nos. 546,650 and 644,086, supra. A pSP64 recombinant clone containing the nucleotide sequence depicted in Table 1, designated as pSP64-VIII, is on deposit at the American Type Culture Collection under Accession Number ATCC 39812.

Restriction endonucleases are used to obtain cleavage of the human factor VIII:C cDNA, hereinafter the DNA source sequence, at appropriate sites in the nucleotide sequence. Unless otherwise noted, restriction endonucleases are utilized under the conditions and in the manner recommended by their commercial suppliers. The restriction endonucleases selected herein are those which will enable one to excise with substantial specificity sequences that code for the portion of the factor VIII:C molecule desired to be excised. BamHI and SacI are particularly useful endonucleases. However, the skilled artisan will be able to utilize other restriction endonucleases chosen by conventional selection methods. The number of nucleotides deleted may vary but care should be taken to insure that the reading frame of the ultimate cDNA sequence will not be affected.

The resulting DNA fragments are then purified using conventional techniques such as those set forth in Maniatis et al., Molecular Cloning, A Laboratory Manual (Cold Spring Harbor Laboratory 1982) the disclosure of which is incorporated herein by reference, and Proc. Natl. Acad. Sci. 76:615-619 (1979). The purified DNA is then ligated to form the sequence encoding the polypeptide of the preferred invention. When necessary or desirable, the ligation may be within an oligonucleotide that resects the cut DNA and maintains the correct translational reading frame using standard ligation conditions. Ligation reactions are carried on as described by Maniatis et al., supra at 2453-6 using the buffer described at page 246 thereof and using a DNA concentration of 1-100 ug/ml, at a temperature of 23°C for blunt ended DNA and 16°C for "sticky ended" DNA. The following double-stranded oligonucleotide is useful when there is BamHI/-SacI deletion such as described infra,

5' P-CATGGACCG-3'

3-TCGAGTACCTGGCCTAG 5';

but other oligonucleotides can be selected by the skilled artisan depending upon the deletions made and reaction conditions.

The DNA sequences encoding the novel procoagulant polypeptides can, in addition to other methods, be derived from the sequence of human factor VIII:C DNA by application of oligonucleotide-mediated deletion mutagenesis, often referred to as "loopout" mutagenesis, as described for example in Morinaga, Y. et al. Biotechnology, 2: 636-639 (1984).

The new DNA sequences containing the various deletions can then be introduced into appropriate vectors for expression in mammalian cells. The procoagulant activity produced by the transiently transfected or stably transformed host cells may be measured by using standard assays for blood plasma samples.

The eukaryotic cell expression vectors described herein may be synthesized by techniques well known to those skilled in this art. The components of the vectors such as the bacterial replicons, selection genes, enhancers, promoters, and the like may be obtained from natural sources or synthesized by known procedures. See Kaufman et al., J. Mol. Biol., 159: 51-521 (1982); Kaufman, Proc. Natl. Acad. Sci. 82: 689-693 (1985).

Established cell lines, including transformed cell lines, are suitable as hosts. Normal diploid cells, cell strains derived from in vitro culture of primary tissue, as well as primary explants (including relatively undifferentiated cells such as haematopoeitic stem cells) are also suitable. Candidate cells need not be genotypically deficient in the selection gene so long as the selection gene is dominantly acting.

The host cells preferably will be established mammalian cell lines. For stable integration of the vector DNA into chromosomal DNA, and for subsequent amplification of the integrated vector DNA, CHO (Chinese hamster ovary) cells are presently preferred. See U.S. Patent 4,399,216. Alternatively, the vector DNA could include all or parts of the bovine papilloma virus genome (Lusky et al., Cell, 36: 391-401 (9184) and be carried in cell lines such as C127 mouse cells as a stable episomal element. Other usable mammalian cell lines include HeLa, COS-1 monkey cells, melanoma cell lines such as Bowes cells, mouse L-929 cells, 3T3 lines derived from Swiss, Balb-c or NIH mice, BHK or HaK hamster cells lines and the like.

Stable transformants then are screened for expression of the procoagulant product by standard immunological or enzymatic assays. The presence of the DNA encoding the procoagulant proteins may be detected by standard procedures such as Southern blotting. Transient expression of the procoagulant genes during the several days after introduction of the expression

vector DNA into suitable host cells such as COS-1 monkey cells is measured without selection by enzymatic or immunologic assay of the proteins in the culture medium.

The invention will be further understood with reference to the following illustrative embodiments, which are purely exemplary, and should not be taken as limiting the true scope of the present invention, as described in the claims.

EXAMPLE 1

10 ug. of the plasmid pACE, a pSP64 (Promega Biotec, Madison, Wis.) derivative, containing nucleotides 562-7269 of human factor VIII:C cDNA (nucleotide 1 is the A of the ATG initiator methionine codon) was subjected to partial BamHI digestion in 100ul containing 50mM Tris.HCl ph 8.0, 50mM $MgCl_2$, and 2.4 units BamHI (New England Biolabs) for 30 minutes at 37°C. The reaction was terminated by the addition of EDTA to 20mM and then extracted once with phenol, once with chloroform , ethanol precipitated and pelleted by centrifugation. DNA was redissolved, cleaved to completion in 50ul using 40 units SacI for 1.5 hours at 37°C. DNA was then electrophoresed through a buffered 0.6% agarose gel. An 8.1 kb fragment corresponding to the partial BamHI-SacI fragment of pACE lacking only the sequence corresponding to nucleotides 2992-4774 of the factor VIII:C sequence was purified from the gel using the glass powder technique described in Proc. Nat. Acad. Sci. 76; 615-619 (1979). Purified DNA was ligated with 100 pmoles of the following double-stranded oligonucleotide

5'P-CATGGACCG-3'

3'-TCGAGTACCTGGCCTAG 5'

using standard ligation conditions. The DNA sequence removed represents the deletion of 584 amino acid sequence beginning with amino acid 998 and continuing through 1581. The oligonucleotide inserted, however, encodes amino acids corresponding to 998-1000. Therefore, the polypeptide encoded contains deletion of 581 amino acids.

DNA was then used to transform competent E. coli bacteria, and DNA from several ampicillin resistant transformants was analyzed by restriction mapping to identify a plasmid harboring the desired SacI-BamHI deletion mutant. DNA from this plasmid was digested to completion with KpnI, which cleaves the plasmid uniquely at nucleotide 1816 of the factor VIII:C coding se-

quence. This DNA was ligated with a KpnI DNA fragment containing nucleotides 1-1815 of factor VIII:C DNA and a synthetic SalI site at nucleotides -11 to -5 and then used to transform competent E. coli bacteria.

Plasmid DNA was isolated and oriented by restriction mapping to identify a plasmid, pBSdK, containing the correct 5' to 3' orientation of the KpnI insert. SalI digestion, which excises the entire polypeptide coding region from the plasmid, was performed and the DNA electrophoresed through a buffered 0.6% agarose gel. The 5.3Kb SalI fragment was purified from the gel as described above. This DNA fragment was ligated with XhoI cut pXMT2 DNA to give rise to plasmid pDGR-2. pXMT2 is a plasmid capable of expressing heterologous genes when introduced into mammalian cells such as the COS-1 African Green Monkey kidney cell line, and is a derivative of the expression vectors described in Kaufman, supra at 689-93. The expression elements are the same as described for plasmid pQ2 except that it contains a deletion of the adenovirus major late promoter extending from -45 to +156 with respect to the transcription start site of the adenovirus major late promoter. mRNA expression in pXMT is driven by the SV40 late promoter. The bacterial replicon, however, has been substituted to render bacteria containing the vector resistant to ampicillin rather than tetracycline. pXMT2 contains a unique Xho I site at a position which allows for expression of inserted cDNA from the SV40 late promoter. This Xho I site is convenient for inserting factor VIII:C cDNA constructs since these are flanked by SalI sites.

Restriction mapping of transformants identified a plasmid, pDGR-2, containing the correct 5' to 3' orientation of the polypeptide coding sequence relative to the direction of transcription from the SV40 late promoter. pDGR-2 is on deposit at the American Type Culture Collection under Accession number 53100.

EXAMPLE 2

Other novel procoagulant proteins may be obtained from constructs produced by oligonucleotide mediated deletion mutagenesis, using for example the "loopout" mutagenesis techniques as described in Morinaga et al., supra. The deletion mutagenesis is performed using expression plasmid pDGR-2 or any other appropriate plasmid or bacteriophage vector. Other methods for oligonucleotide mediated mutagenesis employing single stranded DNA produced with M13 vectors and the like are also suitable. See Zoller et al., Nucl. Acids Res. 10: 6487-6500 (1982). For example, these deletions can be produced using the oligonucleotides

(A) 5' AAAAGCAATTTAATGCCACCCCACCAGTCTTGAAACGCCA

(B) 5' AAAAGCAATTTAATGCCACCGAAGATTTTGACATTTATGA

to cause deletions in factor VIII:C cDNA from nucleotides (A) 2334 to 4974 or (B) 2334 to 5079. The proteins encoded by these constructs contain deletions of (A) 880 and (B) 915 amino acids relative to Factor VIII:C.

The deleted constructs are tested directly, or after subcloning into appropriate expression vectors, in order to determine if the novel proteins possess procoagulant activity. Procoagulant activity was assayed as described in Examples 3 and 4.

EXAMPLE 3

Expression of Procoagulant Molecules in COS Monkey Cells

The expression plasmids containing the modified cDNA's prepared as in Examples 1 or 2 and the full-length cDNA, pXMT-VIII, were introduced into COS-1 cells via the DEAE-dextran transfection protocol. Sompayrac and Dana 1981, Proc. Natl. Acad. Sci. 78: 7575-7578. Conditioned media was harvested 48 hours post-transfection and assayed for factor VIII-type activity as described in Toole et. al., 1984, Nature 312:342-347. The

results of the experiment are summarized in Table 3. Both plasmids containing the modified cDNAs yielded procoagulant activity and, moreover, the activity was greater than that obtained using wild type cDNA. From these data it was concluded that removal of up to 880 amino acids (95,000 daltons) in a defined domain of human factor VIII does not destroy cofactor activity. Furthermore, these abridged procoagulant proteins retain their ability to be activated by thrombin.

TO220X

TABLE 3: EXPRESSION OF ABRIDGED FACTOR VIII MOLECULES

plasmid	# amino acids deleted	chromogenic activity ($mUml^{-1}$)	Clotek activity -IIa	Clotek activity +IIa (fold)
No DNA	-	0		
pXMT-VIII	-	15:1	-	450
pDGR-2	581	114	250	5750 (23X)
pLA-2	880	162	330	9240 (28X)

PS

The plasmids indicated were transfected into COS cells and 48 hr. post-transfection the conditioned media taken for assay by the Kabi Coatest factor VIII:C method (chromogenic activity) and by the one-stage activated partial thromboplastin time (APTT) coagulation assay (Clotek activity) using factor VIII:C deficient plasma as described (Toole, Nature 1984). For thrombin (IIa) activation, samples were pretreated 1-10 min, with 0.2 units/ml thrombin (IIa) at room temperature. Activation coefficients are provided in parentheses. Activity from media from the wild-type (pXMT-VIII) transfection was too low to directly measure Clotek activity before thrombin activation. From other experiments where the wild type factor VIII activity was concentrated, it was demonstrated to be approximately 30-fold activatable.

EXAMPLE 4

Expression of Procoagulant Molecules in CHO Cells

(A) Expression of pDGR-2

The procoagulant expression vector containing a deletion (relative to the Factor VIII:C cDNA) of 581 amino acids (pDGR-2) was transfected with plasmid pAdD26SV(A)#3 (10 ug pDGR-2:1 ug pAdD26SV(A)#3) by $CaPO_4$ coprecipitation into CHO DHFR deficient cells (DUKX-B11) and transformants isolated and grown in increasing concentrations of MTX as described by Kaufman et. al., (1985). One transformant designated J1 exhibited the following activities as a function of resistance to increasing concentrations of MTX.

uM MTX	mUnits/ml/day/10^6 cells*
0	1.46
0.02	322
0.1	499

(B) Expression of pLA-2

The procoagulant expression vector containing a deletion of 880 amino acids (pLA-2) was introduced into CHO DHFR deficient cells (DUKX-B11, Chasin and Urlaub, PNAS 77: 4216-4220, 1980 by protoplast fusion as described (Sandri-Goldin et. al., Mol. Cell. Biol. 1: 743-752). After fusion, fresh medium containing 100 ug/ml of kanamycin, and 10 ug/ml of each of thymidine, adenosine, deoxyadenosine, penicillin, and streptomycin and 10% dialyzed fetal calf serum was added to each plate. The kanamycin was included to prevent the growth of any bacteria which had escaped conversion to protoplasts. Four days later the cells were subcultured 1:15 into alpha-media with 10% dialyzed fetal calf serum, penicillin, and streptomycin, but lacking the nucleosides. Colonies appeared after 10-12 days after subculturing cells into selective media. A group of

transformants were pooled and grown in sequentially increasing concentrations of MTX starting at 0.02 uM with steps to 0.1, 0.2, and 1.0 uM MTX. (LA 3-5 cells; ATCC No. CRL 10101) Results of factor VIII-type activity in cells resistant to increasing concentrations of MTX is shown below.

T0240X

uM MTX	mUnits/ml/day/10^6cells*
0	16
0.02	530
0.2	1170
1.0	1890

* Factor VIII activity was determined by the Kabi Coatest factor VIII:C method (chromogenic activity).